I hold in fact: (1) that small portions of space are of a nature analogous to little hills on a surface which is on the average flat. (2) That this property of being curved or distorted is continually passed on from one portion of space to another after the manner of a wave. (3) That this variation of curvature of space is really what happens in that phenomenon which we call the motion of matter whether ponderable or ethereal. (4) That in this physical world nothing else takes place but this variation, subject, possibly, to the law of continuity.

- *W. K. Clifford (1870)*

Chapter 2

Background

Image processing is a very broad field in computer science. Different practitioners use basic terminology to mean different things, depending on the application. This chapter describes images and image processing techniques, and it provides some notation and vocabulary for this dissertation.

2.1. Notation

Throughout this dissertation I will use a consistent notation for multivalued vs. scalar quantities. Multivalued quantities, vectors, and tensors will be described using bold type (i.e., a bold **p** may be used to represent the Cartesian coordinates (x, y)). I am using row vectors, rather than column vectors unless otherwise specified. Sets will be notated by upper case italics (e.g., $A \subset B$ states that the set A is a subset of set B). A tilde over a character (e.g., $\tilde{u}$) signifies that the character represents a random variable. The symbol $\otimes$ represents the convolution operation, and the $*$ symbol represents a tensor product. A complete list of symbols used in this dissertation is given after the table of contents.

2.2. Images

The term "image" carries many connotations of varying complexity. In general an *image* is a representation of some *scene*, often a view of an object or collection of objects found in nature. Scenes may also be synthetic or computer generated (computer graphics) or simple mathematical functions. Images attempt to capture the structure, information, and features of the scene, representing them as hues, intensities, and other spectral representation. Consider the following definition:

> *Definition:* An *image* is a continuous mapping $I : \mathsf{R}^n \rightarrow \mathsf{R}$. The domain of I, represented by a the set of all points **p** such that $\mathbf{p} \in \mathsf{R}^n$, is an n-dimensional space normally described using a Cartesian coordinate system. The range of this mapping is called the *intensity* or sometimes the "luminance" of I. Images are alternately represented as functions I(**p**).

The intensity I is a single scalar value by this definition. Throughout most of this dissertation, only images with scalar values will be considered. However, the principles presented here are designed with the generalization to multivalued images in mind.

Multivalued images include color images, vector valued images, and images with multiple unnormalized values per location **p**. Chapters 4 and 5 include sections that suggest extensions of this work to multiple spatial dimensions and multivalued images. Projections and extensions to future work involving multivalued images are found in Chapter 6.

2.2.1. Images as a 2D manifold in n-space

When considering the geometry of images, it is often helpful to visualize the mapping from **p** to I as a height field above the n-dimensional spatial domain. For example, consider the 2-dimensional image in its three forms in Figure 2.1. Figure 2.1a shows an image as a collection of intensity values represented by varying levels of grey on a rectilinear grid. Figure 2.1b. shows the same mapping with the intensity as a height field above the x-y plane forming a 2-manifold in 3-space. This representation of images is often called the *intensity surface*. The third view is an iso-intensity contour plot of the image, showing *isophotes,* loci of equal intensity, as closed curves in the plane.

Figure 2.1. Three representations of an image. From left to right: (a) greyscale representation, (b) intensity surface, and (c) isophotes

There has been a great deal of research based on interpreting intensity as height. Pizer, Cullip and Fredericksen use reverse gravity watershed methods for segmentation based on a peak flow model [Pizer 1990]. Griffin, Colchester, and Robinson described segmentation based upon “slope districts,” dividing image regions using to a watershed flow network of graph connected local maxima, minima, and saddle points [Griffin 1991]. There has also been a great deal of focus on image segmentation based on the first order principal directions of gradient and isophote tangent. Early work in segmentation models based on *height ridges* was pioneered by Koenderink and van Doorn [1987]. Approaches describing figural shape based on medial aspects of image structure have been explored by Morse, Pizer, and Liu [Morse 1993]. Their work includes linear, semi-linear, and nonlinear operators across a range of scales that extract medial responses from an image, simultaneously capturing the middles and the widths of objects within the image. Ridges of the subsequent scale-space height field of medialness are then used to characterize structure within the image. Eberly most recently provided a sound mathematical foundation for ridge based analysis by clearly defining height ridges and exploring their robustness [Eberly 1994].

2.2.2. Digital Images

Throughout this dissertation we will consider the realm of *digital images*. When the term "image" is used, it will normally refer to digital representations found in computers. Digital images are collections of computer values addressable using a rectilinear grid.

> *Definition:* A *digital image* is a mapping from a computer addressable domain P to a gamut of representable colors, frequencies, or intensities L. Formally, I is a discrete mapping $\mathrm{I} : P \to L$; $P \subset \mathbb{N}^n$, $L \subset \mathbb{R}$ where $\mathbb{N}$ represents the natural numbers (non-negative integers). P is limited to a finite subset of $\mathbb{N}^n$. L is limited to rational numbers of finite precision. An individual location $\mathbf{p} \in P$ along with its intensity value $\mathrm{I}(\mathbf{p})$ is called a "picture element" or more commonly a *pixel*.

The grid may be of any dimension, but this work in greater part only deals with images of two dimensions (i.e., the domain P is the set of all points $\mathbf{p} \in P$ that may be represented in the image such that P is a finite, compact subset of $\mathbb{N}^2$). A Cartesian coordinate system is used to specify the individual location of each pixel. Image intensity L is depicted using the precision of the available digital representation and discretely addressed through the grid array. Figure 2.2 shows the same image function in Figure 2.1 as a digital image with a raster resolutions of 64×64 pixels. As with most digital image representations, the figure uses a left-handed coordinate system. The y-axis points toward the bottom of the page.

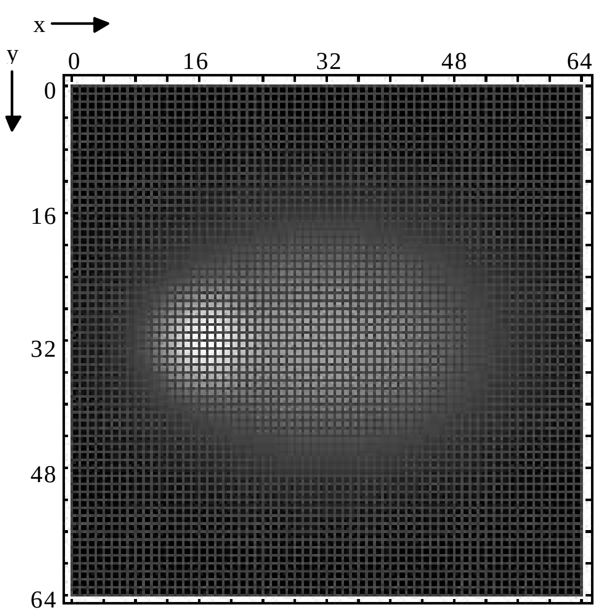

Figure 2.2. The image in Figure 2.1 represented as a digital image with a raster resolution of 64×64 pixels.

2.3. Invariance

The sampling and discretization processes used to create digital images often impede image analysis. What is desired is the analysis of images in such a manner that any measurements are made relative to the scene of which the image is only a partial representation regardless of the sampling process. One may consider a digital image to be a grated window through which a much wider, continuous scene is viewed. While it is impossible to escape the constraints of sampling and the finite size of the window, those constraints should have limited bearing on the understanding of the scene beyond the window. The relative placement, orientation, and size of the window frame as well as the internal spacing, orientation and width of the grid within the grating have no relevance to distances, sizes, luminance and structure of the scene beyond.

When processing an image, it is tempting to employ the implicit coordinate system of the digital representation. However, its use imposes an artificial structure upon the image. Since a digital image is a discrete sampling of a subset of an infinite continuous scene, any set of basis vectors used to describe the digital image that is not somehow related to the original scene creates arbitrary metrics by which the discrete information is evaluated. Care must be taken at all times not to let these metrics influence the image analysis methods since repositioning or resampling of the scene may alter the metrics, thus possibly undesirably altering the outcome of the analysis *without an alteration of the underlying scene*.

This observation can be restated in more formal terms. Methods for defining objects within images should be *invariant* with respect to translation, rotation and zoom. The following presentation of invariance, though not original, was presented by Eberly [Eberly 1993]. For elaboration see Olver, Sapiro and Tannenbaum [Olver 1994]. Consider the following definitions.

> *Definition:* A *vector field* on $\mathbb{R}^n$ is a map $\vec{\mathbf{V}}: X \subset \mathbb{R}^n \to \mathbb{R}^n$ that assigns to each point $\mathbf{x} \in X$ a vector $\vec{\mathbf{V}}(\mathbf{x}) \in \mathbb{R}^n$. An example of a vector field is the gradient field of a scalar function.
>
> *Definition:* Given a vector field $\vec{\mathbf{V}}: \mathbb{R}^n \to \mathbb{R}^n$, a *flow line* is a curve $\mathbf{s}(t)$ such that for any value of t_0, the tangent to $\mathbf{s}$ at $\mathbf{s}(t_0)$ is equal to $\vec{\mathbf{V}}(\mathbf{s}(t_0))$. That is, for any t, $\mathbf{s}(t)$ is everywhere tangent to $\vec{\mathbf{V}}$.
>
> *Definition:* Given a vector field $\vec{\mathbf{V}}: X \subset \mathbb{R}^n \to \mathbb{R}^n$, an *invariant of* $\vec{\mathbf{V}}$ is a function f: $\mathbb{R}^n \to \mathbb{R}$ such that its directional derivative satisfies $\vec{\mathbf{V}} \bullet \nabla f = 0$. That is, given any flow line $\mathbf{s}(t)$ of $\vec{\mathbf{V}}$ and $t_0 \neq t_1$, $f(\mathbf{s}(t_0)) = f(\mathbf{s}(t_1))$. Informally, the value of the invariant f remains constant along the direction of the field $\vec{\mathbf{V}}$.

A typical manipulation of an image involves some transformation of the coordinate system used to describe the image. An operation such as translation or rotation is therefore a map $\mathbf{T}: \mathbb{R}^n \to \mathbb{R}^n$ of an image I onto I′. That is, the domain of the coordinate transform $\mathbf{T}$ is the space of images $I : X \subset \mathbb{R}^n \to \mathbb{R}$, and its range is the space of images $I' : X' \subset \mathbb{R}^n \to \mathbb{R}$, where the basis of the X is $(\mathbf{i}, \mathbf{j}, \mathbf{k})$ and the basis of X' is (**i**′, **j**′, **k**′). So-called *invariant theory* is the subdiscipline of tensor calculus that deals with functions that remain constant under these change of bases or coordinate transforms or both [ter Haar Romeny 1991ab].

If the coordinate transformation is parameterized, the result is a mapping of the image onto an infinity of possible resultant images, each depending on the magnitude of the parameters. For example, the rotation of an image $I : X \subset \mathbb{R}^n \rightarrow \mathbb{R}$ may be parameterized on θ, the angle of rotation. The result is called a *prolongation* of the image space (X,I) along a vector field, represented by $pr(\mathbf{x}, I, \theta)$ where $\mathbf{x} \in X$. The result is a map of the n-dimensional scalar image I onto an (n+1)-dimensional space, with the operation (rotation through the angle θ in this example) generating a vector field on the n+1 dimensional space. Given some value $\mathbf{x}_0 \in X$, $pr(\mathbf{x}_0, I, \theta)$ is a flow line of the rotating vector field passing through the point $pr(\mathbf{x}_0, I, 0)$.

The study of invariants in image processing is thus described as the analysis of functions that remain constant under a continuous change of bases, through some coordinate transformation or deformation. An image invariant with respect to some operation described by vector field $\vec{\mathbf{V}}$, is a function f such that given two values $t_0 \neq t_1$, $f(pr(\mathbf{x}, I, t_0)) = f(pr(\mathbf{x}, I, t_1))$ [Olver 1994].

The study of Lie groups is the analysis of functions that are invariant under particular sets of operations, the characteristics of those operations, and the equivalence classes among the groups of invariant functions. A presentation of the algebra of Lie groups is beyond the scope of this dissertation. For this dissertation, it will be enough to show that particular functions are invariant with respect to particular vector fields (i.e., for particular operations).

Lie group theory applies naturally to differential scale space invariants. For a description of scale space differential invariants and their associated mathematics see Olver [Olver 1993].

2.3.1. Gauge Coordinates

Invariance with respect to spatial rotation is important in many differential geometric measurements. One way of achieving rotational invariance is to measure image geometry relative to a coordinate system, created at each pixel, that is based on the local structure of the image. Such an approach frees differential geometric measurements from the global Cartesian coordinate system. Taken together, the isophote tangent and the gradient vector form a coordinate frame with corresponding distance metrics at each point **p** of the image provided that particular conditions hold. These coordinates are described as the *gauge coordinates* by Koenderink for surfaces [Koenderink 1984] and later by ter Haar Romeny for images [ter Haar Romeny 1991ab].

The existence of the gauge coordinates depends on the *gauge condition*, the property that the basis vectors exist and are unique. While there are many locations where the gauge condition is not met, they are countably infinite and represent extrema within the image. For generic images, the gauge condition can be assumed for most points within the image. Pathological cases represent non-generic images where the extrema are uncountably infinite.

The gauge directions are orthogonal. The relative magnitude of each of these direction vectors implies a distance metric. The magnitudes of the gauge are invariant with respect to rotation and translation of the image. This natural coordinate frame provides an orientation and distance metric for every point within the image that can be used to make local geometric measurements.

2.4. Scale Space

In the analysis of images, measurements are made of image features including spatial distance, intensity differences, and the relationship between space and intensity. Each of these measurements requires that some tolerance, aperture, or *scale* be used as the basis of the measurement. It is illuminating to treat the image not only at a single scale but rather to make measurements of the image over a range of scales. Multiscale analysis yields insights into both fine and coarse structures.

The discrete nature of digital images necessitates the use of some aperture or scale function. If an analytic form is designated for the scene description, an infinitesimal scale may be specified yielding instantaneous measures of image geometry. However, when processing digital images, the analytic form of the underlying image function is often not known. Discrete, evenly spaced samples make instantaneous measures at infinitesimal scale impossible. Some distance metric must be applied and some measurement tolerance assumed.

It is tempting to use the size of a single pixel as a unit length and the Cartesian coordinates supplied by the pixel grid as the coordinate frame for image analysis. If one recognizes that aside from convenience there is no compelling reason to select that spacing or those grid directions as the elementary distance metric or geometric basis for the image, the tolerance, aperture, or metric with which distances are measured becomes a free parameter in the evaluation of images.

Let the symbol $\otimes$ represent the convolution operation expressed by

$$f \otimes g = f(x) \otimes g(x) = \int_{-\infty}^{\infty} f(x - \tau) g(\tau) d\tau \tag{2-1}$$

The convolution operation is linear and shift invariant. Convolution can be interpreted as a weighted average of neighborhoods in an image in which the neighborhood is defined by some filter kernel. This essential spatial aspect, added to image measurement operations, means that measurement is dependent on the size or scale of the convolution kernel. This process introduces scale as a parameter to image measurement. Consider the following definition.

> *Definition:* A *scale-space image* is a continuous mapping I : $S \rightarrow \mathbb{R}$. *S* is the set of all points (**p**, σ) where $\mathbf{p} \in \mathbb{R}^n$, and the additional dimension $\sigma \in \mathbb{R}^+$ is attributed to the aperture with which intensity measurements are made. The range values of this mapping I(**p**,σ) are intensity measurements made of the scene at point **p** at scale σ. The range of the mapping is called the *intensity of I at scale s*. Scale space images are alternately represented as function of the vector valued point **p** and scale σ, such that $I(\mathbf{p}, \sigma) = I(\mathbf{p}) \otimes H(\sigma, \mathbf{p})$ where H is the scaling filter kernel and ⊗ is the convolution operation.

Scale-space representations are distinct from other *coarse-to-fine* representations. Coarse-to-fine systems involve the reduction of the raster resolution of the digital image as the image space moves through representations from fine to coarse detail. The wavelet transform and its related work [Mallat 1989] as well as the work of Burt and Adelson on the Laplacian pyramid for image compaction [Burt 1986] are examples of coarse-to-fine techniques.

The above definition of scale-space images above does not describe the selection of the scaling filter kernel nor the effect that the scale parameter σ has on the shape or size of the kernel. Work by previous researchers has systematically introduced requirements and constraints on the scaling filter kernel, refining the definition of scale space.

2.4.1. Differentiation

It has been repeatedly demonstrated that the process of differentiation is mildly ill-posed. At the turn of the century, Hadamard defined an operation to be well posed under the following conditions: (1) its solution exists, (2) the solution is unique, and (3) the solution depends continuously on the initial data. This question of ill-posedness has as a consequence that differentiation "lacks the necessary robustness for practical implementations." In particular, "small perturbations of the input data may have an arbitrarily large effect on the measured derivative values" [Florack 1993].

Torre and Poggio present a well posed differentiation operation based upon the convolution of a linear filter kernel with the input data or function. Paraphrasing Torre and Poggio, the process of numerical differentiation may be regularized by first convolving the data with a Gaussian or a "Gaussian-like" filter kernel before differentiation. They indicate, however, that the two processes of differentiation and regularization are separate. In particular, they suggest that filtering for regularization should be a preprocessing step while filtering for multiresolution analysis should be performed after the differentiation operation when nonlinear differential operators are used (e.g., the second derivative in the gradient direction $\partial^2/\partial v^2$ is a nonlinear differential operator) [Torre 1986].

Ter Haar Romeny and Florack observe that the two operations may be combined. They point out that the operations of differentiation and convolution with a linear operator are commutative and associative. That is,

$$\frac{\partial^n}{\partial x^n} I \otimes H = \frac{\partial^n}{\partial x^n} H \otimes I \tag{2-2}$$

where I is the image, and H is a filter kernel. They choose the Gaussian as their kernel. The resulting 2-D differential operators through the 4th order are shown as images in Figure 2.3.

The result is that differentiation may be regularized as the image is filtered for multiresolution analysis. This combined differentiation and filtering leads to an elegant description of multiscale image analysis based on the axiomatic principles of differential geometry [ter Haar Romeny 1991a]. In his dissertation, Luc Florack details the mathematical structure of this operation, citing Schwarz's work published in the 1950s on *regular tempered distributions*. Florack presents the details of this theory as applied to images and the construction of linear scale spaces [Florack 1993].

Florack points out that the *smoothness* of the input function is not relevant to the condition of ill-posedness in differentiation. The condition of being either well- or ill-posed is an artifact of the operation itself and not the space of input functions. What is required is a regularization of the *process* of differentiation and not the data upon which the operation is performed. Thus while the two approaches may achieve similar results, the multiresolution image analysis methods of ter Haar Romeny and Florack are axiomatically superior to the two or three stage processes (regularize, differentiate, filter) of Torre *et al.*

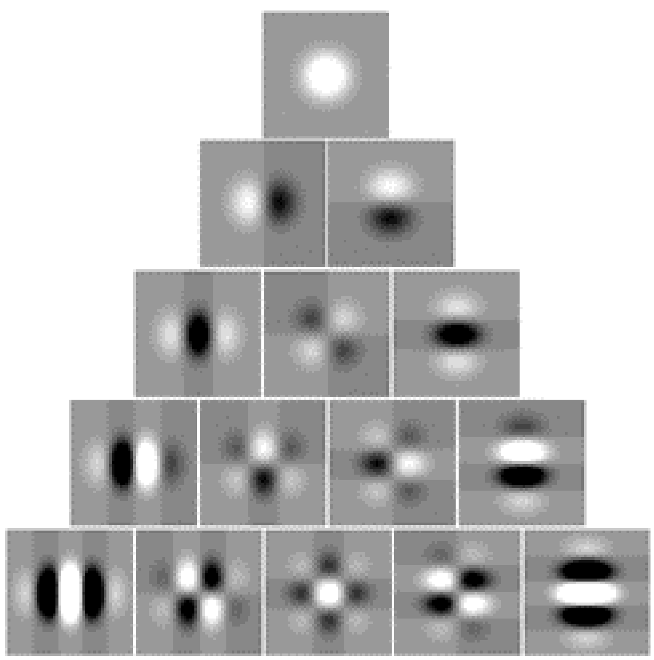

Figure 2.3. 2-D Gaussian derivative filter kernels through the 4th order.

2.4.2. The Gaussian as a unique Regular Tempered Distribution

The Gaussian has been identified as a unique operator in the generation of scale space. Pioneering work by Witkin [Witkin 1983], Koenderink [Koenderink 1984], Babaud *et al.*

[Babaud 1986], and Yuille and Poggio [Yuille 1986] all point toward the Gaussian as the "unambiguous" choice for a regularizing operator for the generation of linear scale spaces [ter Haar Romeny 1991b]. A more complete discussion of the Gaussian providing a regular tempered distribution is provided by Florack [Florack 1993, 1994b].

An analysis of the selection of the Gaussian as the scale space filter is beyond the scope of this dissertation. However, a list of its properties and a brief discussion of their importance is included. A multivariate Gaussian is a function of the form

$$G(\Sigma, \mu, \mathbf{p}) = A\, e^{-\frac{1}{2}\left((\mathbf{p}-\mu)\Sigma^{-1}(\mathbf{p}-\mu)^T\right)} \tag{2-3}$$

where n is the dimension of the kernel or image space, Σ is the $n \times n$ covariance matrix (of rank n) for the Gaussian spatial distribution, and A is a real scaling constant. If one specifies a scalar value σ that represents the aperture size or *scale* of the Gaussian operator and constrains $\Sigma = \sigma I$, where I is the $n \times n$ identity matrix, and sets $\mu = 0$, then $A = \frac{1}{(2\pi\sigma)^{n/2}}$, and equation (2-3) is reformed as

$$G(\sigma, \mathbf{p}) = \frac{1}{(2\pi\sigma)^{n/2}}\, e^{-\frac{1}{2}\left(\frac{(\mathbf{p}\bullet\mathbf{p})}{\sigma}\right)} \tag{2-4}$$

The resulting filter kernel $G(\sigma, \mathbf{p})$ has the following properties:

Rotational invariance: The Gaussian function in equation (2-4) is isotropic. When used as a filter kernel, it has no preferred direction or orientation.

Solutions to the linear diffusion equation: The Gaussian and its spatial derivatives are solutions to the heat equation (the linear diffusion equation) given by the function

$$\frac{\partial I}{\partial t} = \nabla \bullet c\nabla I \tag{2-5}$$

In their explorations of scale space, Eberly and ter Haar Romeny both choose $t = \sigma^2 / 2$ and $c = 1$, generating (2-4) as the corresponding solution.

Generates no spurious resolution: Gaussian blurring, based as it is on the linear diffusion equation, is *causal*; no new level curves appear in scale space that do not originate at some smaller aperture. That is, diffusion obeys the maximum principle. Simply stated, while new local extrema or other critical points can be created through Gaussian blurring, the number of them can be guaranteed to decrease in the limit with increasing scale.

Normalization: This particular Gaussian kernel is normalized to unit weight. Convolving with this Gaussian conserves the overall intensity of the input signal. The Gaussian retains this normalization through changes in the scale parameter. While not an essential property, this is a useful attribute of a filter kernel in practice.

Separable: The Gaussian is linear and separable along Cartesian coordinates. This is mainly a matter of mathematical and computational convenience. This property enables a simplification of the numerics of scale space computation, making tractable the analysis of images in response to changing scale.

Central limit theorem: The central limit theorem specifies the Gaussian as the aggregate kernel of k applications of any linear shift invariant operators with positive kernels as k approaches infinity. That is, repeated filtering with any arbitrary set of positive linear filter kernels eventually approximates blurring with a Gaussian.

2.4.3. Zoom Invariance

Adding scale as a free parameter for image analysis yields an n + 1 dimensional space to describe a single n dimensional image. Work by Florack and ter Haar Romeny suggests that invariance of image measurements with respect to changes of scale is also desirable. They suggest a distance metric that preserves the Euclidean nature of scale space [ter Haar Romeny 1991ab]. Eberly extended this concept of *zoom-invariance* where measurements remain invariant under changes of uniform magnification of both space and scale. The resulting metric for measuring distances in a Gaussian (linear) scale space is

$$\partial\phi^2 = \frac{\partial\mathbf{p} \bullet \partial\mathbf{p}}{\sigma^2} \quad + \quad \rho^2 \frac{\partial\sigma^2}{\sigma^2} \tag{2-6}$$

where ρ is a constant relating rate of change in the scale dimension to the rate of change in image space. This suggests a normalization of scale space measurements in which *all spatial differences and scale differences are made relative to the scale at which the difference is measured.* Eberly’s dimensionless 1-form used in scale space measurements is

$$\frac{\partial\mathbf{p}}{\sigma} \qquad \rho\frac{\partial\sigma}{\sigma} \tag{2-7}$$

This metric describes a corresponding non-Euclidean hyperbolic space. A complete discussion is available in his dissertation as well as in book chapter form [Eberly 1994ab]. Certain consequences of using this metric are explored in Chapter 3.

2.4.4. Gaussian Scale Space

Given the discussion above and the work of previous researchers, it is possible to refine the earlier definition of scale space and specify both the structure of scale space and the causal operators that generate the space. The constraints of linearity, separability, causality, and zoom invariance dictate the use of a Gaussian as the unique scale-space operator.

> *Definition:* A *Gaussian scale-space image* is a continuous mapping I : $S \rightarrow \mathbb{R}$. S is the set of all points ($\mathbf{p}$, σ) where $\mathbf{p} \in \mathbb{R}^n$, and the scale parameter $\sigma \in \mathbb{R}^+$ (i.e., $\sigma \in \mathbb{R}$ and $\sigma > 0$). The range of this mapping is called the *intensity of I at position **p** and scale s*. The aperture with which geometric

measurements are made is an isotropic Gaussian $G(\sigma, \mathbf{p})$. Scale space images are alternately represented as functions of the vector valued point **p** given scale σ, or mathematically as $I(\mathbf{p} \mid \sigma) = I(\mathbf{p}) \otimes G(\sigma, \mathbf{p})$.

2.5. Image Statistics

To this point this discussion has considered the tolerances in spatial measurements. Issues regarding the well-posedness and regularization of differentiation have been addressed by considering the relative scales by which space is measured. However, all physical measurement is subject to some error or noise in the input signal. The sensitivity of analytical methods to variations in the input signal is an important issue when considering the robustness of any algorithm. The purpose of this work is in part to understand and incorporate tolerances of intensity measurement into scale space theory.

The computation of statistics of image space has a long history. Filtering methods based on local neighborhood statistics such as *median filtering* can be found throughout the literature. Image contrast enhancement techniques based on histogram equalization have also been explored and are in use in medical as well as other production environments [Pizer 1987]. There exist numerous methods for performing segmentation and classification of images based on statistical pattern recognition [Duda 1973]. Statistically based relaxation filters founded on the theory of Markov processes (Markov Random Fields) [Geman 1984], and expectation-maximization methods [Dempster 1977] also have a long history. Geiger and Yuille attempt to capture these various methods as well as those of nonlinear diffusion, discussed in Chapters 4 and 5, in a taxonomy of segmentation algorithms. Their treatment serves as a succinct survey of the common threads shared by different segmentation algorithms [Geiger 1991].

This section presents pertinent background to support work presented later in this thesis. It is not intended as an exhaustive treatise on image statistics.

2.5.1. The Normal Density vs. the Gaussian Filter Kernel

The Gaussian function represents one of the most important probability density functions used in the field of statistics as well as a mathematically interesting linear operator in image processing.

For a random variable $\tilde{v}$ with a univariate Normal density, its probability distribution function is completely specified by two parameters, the mean $\mu_{\tilde{v}}$ and variance $\mu_{\tilde{v}}^{(2)}$. The mean is the first moment of the probability density function of $\tilde{v}$, and the variance is the second central moment. They are by definition

$$\begin{aligned} \mu_{\tilde{v}} &= \langle \tilde{v} \rangle \\ \mu_{\tilde{v}}^{(2)} &= \left\langle \left(\tilde{v} - \mu_{\tilde{v}} \right)^2 \right\rangle \end{aligned} \tag{2-9}$$

where the angle brackets $\langle \; \rangle$ represent the expectation operator.

The univariate Normal density function corresponding to the mean $\mu_{\tilde{v}}$ and variance $\mu_{\tilde{v}}^{(2)}$ is defined as $N_{\mu_{\tilde{v}},\mu_{\tilde{v}}^{(2)}}(\tilde{v})$ where

$$N_{\mu_{\tilde{v}},\mu_{\tilde{v}}^{(2)}}(\tilde{v}) = \frac{1}{\sqrt{2\pi\mu_{\tilde{v}}^{(2)}}} e^{-\frac{1}{2}\frac{\left(\tilde{v}-\mu_{\tilde{v}}\right)^2}{\mu_{\tilde{v}}^{(2)}}} \tag{2-10}$$

It is important to distinguish the two uses of this function and to clearly delineate the use of the size parameter that describes the width of the Gaussian shaped filter kernel from the square root of the variance or spread of the probability distribution function. At all times in this dissertation the term *scale* and the Greek letter σ will be used to refer to the width of the Gaussian scale-space operator. The term *standard deviation* will describe the square root of the variance of the probability density function.

2.5.2. Noisy Images

Consider an image with added noise $\tilde{I}(\mathbf{p}) = I(\mathbf{p}) + \tilde{u}(\mathbf{p})$ where $\tilde{u}(\mathbf{p})$ is a zero mean random variable that has a Normal probability density (p.d.f.) with variance v_0, $N_{0,v_0}(\tilde{u}(\mathbf{p}))$. Specifically,

$$\text{p.d.f.}(\tilde{u}(\mathbf{p})) = N_{0,v_0}(\tilde{u}(\mathbf{p})) = \frac{1}{\sqrt{2\pi v_0}} e^{-\frac{1}{2}\frac{\tilde{u}(\mathbf{p})^2}{v_0}} \tag{2-11}$$

If $\tilde{u}(\mathbf{p})$ is uncorrelated as a function of **p**, it is called *white*. When considering white noise, the irrelevant spatial parameter **p** is dropped from the notation (i.e., $\tilde{u}(\mathbf{p})$ becomes $\tilde{u}$). Throughout this dissertation unless otherwise indicated, the term “noisy image” will refer to an image with zero-mean additive Gaussian white noise.

Computing the mean and the variance of $\tilde{I}(\mathbf{p})$ yields

$$\mu_{\tilde{I}}(\mathbf{p}) = \left\langle \tilde{I}(\mathbf{p}) \right\rangle = \langle I(\mathbf{p}) + \tilde{u} \rangle = I(\mathbf{p}) + \langle \tilde{u} \rangle = I(\mathbf{p}) \tag{2-12}$$

$$\mu_{\tilde{I}}^{(2)}(\mathbf{p}) = \left\langle \left(\tilde{I}(\mathbf{p}) - \mu_{\tilde{I}}(\mathbf{p})\right)^2 \right\rangle = \left\langle \left((I(\mathbf{p}) + \tilde{u}) - I(\mathbf{p})\right)^2 \right\rangle = \left\langle \tilde{u}^2 \right\rangle = v_0 \tag{2-13}$$

2.5.3. Statistical Measures as Invariants: Mahalanobis Distances

In measuring intensity values, it is often important to measure the difference between the mean value and the sample value relative to the shape and spread of the distribution. In the univariate case the commonly used metric is

$$r^2 = \left(r(\mathbf{p}_0, \tilde{u})\right)^2 = \frac{\left(\tilde{u}(\mathbf{p}_0) - \mu_{\tilde{u}}\right)^2}{\mu_{\tilde{u}}^{(2)}} \tag{2-14}$$

where $\tilde{u}(\mathbf{p}_0)$ is the observed sample value being measured. The non-negative value r is called the *Mahalanobis distance* from $\tilde{u}(\mathbf{p}_0)$ to $\mu_{\tilde{u}}$.

It is straightforward to show that r is invariant under linear functions of intensity. Let $\tilde{u}$ be subjected to a linear transformation of intensity such that,

$$\tilde{u}' = a\tilde{u} + b \tag{2-15}$$

where a and b are scalar constants. Thus,

$$\tilde{u}'(\mathbf{p}) = a\,\tilde{u}(\mathbf{p}) + b \tag{2-16a}$$

$$r'^2 = \left(r(\mathbf{p}_0, \tilde{u}')\right)^2 = \frac{\left(\tilde{u}'(\mathbf{p}_0) - \mu_{\tilde{u}'}\right)^2}{\mu^{(2)}_{\tilde{u}'}} \tag{2-16b}$$

Invariance with respect to linear functions of intensity implies $r' = r$. This is shown by first computing the mean and variance of $\tilde{u}'$:

$$\begin{aligned}
\mu_{\tilde{u}'} &= \langle \tilde{u}' \rangle = \langle a\tilde{u} + b \rangle = a\langle \tilde{u} \rangle + b = a\mu_{\tilde{u}} + b \\
\mu^{(2)}_{\tilde{u}'} &= \left\langle \left(\tilde{u}' - \mu_{\tilde{u}'}\right)^2 \right\rangle = \left\langle \left(a\tilde{u} + b - \mu_{\tilde{u}'}\right)^2 \right\rangle \\
&= \left\langle \left(a\tilde{u} + b - (a\mu_{\tilde{u}} + b)\right)^2 \right\rangle = a^2 \left\langle \left(\tilde{u} - \mu_{\tilde{u}}\right)^2 \right\rangle = a^2 \mu^{(2)}_{\tilde{u}}
\end{aligned} \tag{2-17}$$

The Mahalanobis distance associated with the transformed random variable is

$$r'^2 = \frac{\left(\tilde{u}'(\mathbf{p}) - \mu_{\tilde{u}'}\right)^2}{\mu^{(2)}_{\tilde{u}'}} = \frac{\left(\left(a\,\tilde{u}(\mathbf{p}) + b\right) - \left(a\,\mu_{\tilde{u}} + b\right)\right)^2}{a^2\,\mu^{(2)}_{\tilde{u}}} = \frac{\left(\tilde{u}(\mathbf{p}) - \mu_{\tilde{u}}\right)^2}{\mu^{(2)}_{\tilde{u}}} \tag{2-18}$$

Thus $r' = r$. ❑

In the case of multivariate distributions, their corresponding multivariate Mahalanobis distances are also invariant to linear functions of the separate random variables. It is this property of intensity invariance that I wish to preserve in the ideas that follow in later chapters.

2.5.4. Calculating Central Moments

Section 2.4.1. describes the Normal density and the use of the first moment, or mean, and the second central moment, or variance, of a random variable. While these values are defined using the expectation operator, it is important to provide a stronger working definition for these values and to generalize this definition to higher order moments.

If the probability density function $f(\tilde{u})$ of a random variable $\tilde{u}$ is known, the mean or first moment is calculated by

$$\mu_{\tilde{u}} = \langle \tilde{u} \rangle = \int_{-\infty}^{\infty} f(\tilde{u})\,\tilde{u}\; d\tilde{u} \tag{2-19}$$

The central moments are

$$\mu_{\tilde{u}}^{(r)} = \left\langle \left(\tilde{u} - \mu_{\tilde{u}}\right)^r \right\rangle = \int_{-\infty}^{\infty} f(\tilde{u}) \left(\tilde{u} - \mu_{\tilde{u}}\right)^r \, d\tilde{u} \tag{2-20}$$

These moments are known as the central moment of $\tilde{u}$ of order r. By the definitions above, $\mu_{\tilde{u}}^{(0)} = \int_{-\infty}^{\infty} f(\tilde{u})\; d\tilde{u} = 1$, and $\mu_{\tilde{u}}^{(1)} = \int_{-\infty}^{\infty} f(\tilde{u})\left(\tilde{u} - \mu_{\tilde{u}}\right) d\tilde{u} = \int_{-\infty}^{\infty} f(\tilde{u})\; \tilde{u}\; d\tilde{u} - \mu_{\tilde{u}} = 0$. If $f(\tilde{u})$ is not known and $\tilde{u}$ is represented by a sample of n values $[u_1, u_2, \ldots u_n]$, the mean may be approximated by a weighted sum of the sample values

$$\mu_{\tilde{u}} \approx \sum_{i=1}^{n} w_i u_i \qquad \text{with weights } [w_1, w_2, \ldots, w_n], \qquad \sum_{i=1}^{n} w_i = 1 \tag{2-21}$$

The central moments of $\tilde{u}$ can be calculated from discrete samples $[u_1, u_2, \ldots u_n]$, by

$$\mu_{\tilde{u}}^{(r)} \approx \sum_{i=1}^{n} w_i (u_i - \mu_{\tilde{u}})^r \quad \text{with weights } [w_1, w_2, \ldots, w_n], \qquad \sum_{i=1}^{n} w_i = 1 \tag{2-22}$$

For the common case of uniformly distributed random variables, $w_1 = w_2 = \ldots = w_n = 1/n$.

Of interest are the first four moments, all of which are named. The first two, the mean and variance have already been discussed. The other two are also named and are often normalized to exhibit invariance with respect to linear functions of intensity.

Skewness – The third central moment can be normalized by dividing it by $\left(\sqrt{\mu_{\tilde{u}}^{(2)}}\right)^3$. The resulting value, called the *skewness* of $\tilde{u}$, describes the asymmetry of the probability distribution function. Skewness is calculated by

$$\gamma_{\tilde{u}}^{(3)} = \frac{\mu_{\tilde{u}}^{(3)}}{\left(\sqrt{\mu_{\tilde{u}}^{(2)}}\right)^3} \tag{2-23}$$

This measure of skewness is invariant with respect to linear functions of intensity.

Kurtosis – The fourth central moment can similarly be normalized by dividing it by $\left(\mu_{\tilde{u}}^{(2)}\right)^2$. The resulting value, called the *kurtosis* of $\tilde{u}$, describes the spread or peakedness of the probability distribution function relative to the Normal density. Kurtosis is calculated by

$$\gamma_{\tilde{u}}^{(4)} = \frac{\mu_{\tilde{u}}^{(4)}}{\left(\mu_{\tilde{u}}^{(2)}\right)^2} \tag{2-24}$$

Kurtosis is also invariant with respect to linear functions of intensity.

The literature distinguishes different conditions of peakedness described by the kurtosis statistic. The kurtosis of the Normal density is 3. Distributions whose kurtosis =

3 are said to be *mesokurtic*. Those with kurtosis > 3 are *platykurtic* or more table-like, and those with kurtosis < 3 are *leptokurtic*.

2.5.5. Characteristic Functions

There exists a transform to convert probability density functions to a corresponding frequency domain. The frequency based representation is commonly named the *characteristic function* of a distribution, and it has several interesting properties. If the probability density function is known in advance, the moments of that function can be calculated from the characteristic function. If all the moments of the density function are known, it can be shown that the collective moments uniquely determine the characteristic function and thereby uniquely determine the density function. The inverse transform from moments to probability distribution is possible only if all moments are finite and if the Maclaurin series expansion used to reconstruct the probability distribution function converges absolutely near the first moment.

Univariate Characteristic Functions

> *Definition:* If $\tilde{u}$ is a random variable and $f(\tilde{u})$ is its probability density function, the *characteristic function* of $\tilde{u}$ is defined to be $F(\omega)$, the Fourier transform (not normalized to 2π) of $f(\tilde{u})$ (see equation (2-25a)). Because it is a Fourier transform, an inverse transform of $F(\omega)$ yields the original probability density function (equation (2-25b)).

$$F(\omega) = \int_{-\infty}^{\infty} f(\tilde{u}) e^{i\omega\tilde{u}} d\tilde{u} \qquad\qquad f(\tilde{u}) = \tfrac{1}{2\pi}\int_{-\infty}^{\infty} F(\omega) e^{-i\omega\tilde{u}} d\omega \tag{2-25a,b}$$

From the moment theorem it is possible to obtain $f(\tilde{u})$ from its moments. Given a series of central moments that describe $f(\tilde{u})$, it is possible to generate the probability distribution using (2-25b) and the following Taylor expansion equation.

$$F(\omega) = \sum_{n=0}^{\infty} \tfrac{1}{n!} \mu_{\tilde{u}}^{(n)} (i\omega)^n \tag{2-26}$$

The value s is often substituted for the complex value $i\omega$, to create the *moment generating function* $\Phi(s)$.

$$\Phi(s) = \int_{-\infty}^{\infty} f(\tilde{u}) e^{s\tilde{u}} d\tilde{u} \tag{2-27}$$

This expression is called the moment generating function because the k-th derivative of this function with respect to s produces the k-th moment of the probability density function.

A Simple Univariate Example (a Gaussian Normal Distribution)

An approximation to a probability density function can be generated by using the first few central moments in the above zero centered Taylor expansion. Certain error terms will be generated, and the inverse characteristic function may require limitation to an integration domain. The inverse integral to calculate the approximate probability distribution function may not exist. Therefore, certain constraints must be imposed on the expansion. That is, the expansion shown in equation (2-26) must converge absolutely at $\omega = 0$.

Consider the probability distribution function $f(\tilde{u})$ to be a Normal distribution with a zero mean and standard deviation of σ. The probability distribution function $f(\tilde{u})$ and its corresponding characteristic function are described by equations (2-28a) and (2-28b) respectively.

$$f(\tilde{u}) = N_{0,v_0}(\tilde{u}) = \frac{1}{\sqrt{2\pi v_0}} e^{-\frac{1}{2}\frac{\tilde{u}^2}{v_0}} \qquad F(\omega) = e^{-\frac{1}{2}v_0(\omega)^2} \tag{2-28a,b}$$

Substituting s for $i\omega$ produces the moment generating function $\Phi(s)$. Differentiating $\Phi(s)$ yields moments of the Gaussian probability distribution. The first six moments are

$$\begin{aligned}
\Phi(s) &= e^{\frac{1}{2}v_0 s^2} \\
\tfrac{\partial}{\partial s}\Phi(s) &= s\, v_0 e^{\frac{1}{2}v_0 s^2} \\
\tfrac{\partial^2}{\partial s^2}\Phi(s) &= v_0 e^{\frac{1}{2}v_0 s^2} + s^2 v_0{}^2 e^{\frac{1}{2}v_0 s^2} \\
\tfrac{\partial^3}{\partial s^3}\Phi(s) &= 3s\, v_0{}^2 e^{\frac{1}{2}v_0 s^2} + s^3 v_0{}^3 e^{\frac{1}{2}v_0 s^2} \\
\tfrac{\partial^4}{\partial s^4}\Phi(s) &= 3v_0{}^2 e^{\frac{1}{2}v_0 s^2} + 6s^2 v_0{}^3 e^{\frac{1}{2}v_0 s^2} + s^4 v_0{}^4 e^{\frac{1}{2}v_0 s^2} \\
\tfrac{\partial^5}{\partial s^5}\Phi(s) &= 15s v_0{}^3 e^{\frac{1}{2}v_0 s^2} + 10s^3 v_0{}^4 e^{\frac{1}{2}v_0 s^2} + s^5 v_0{}^5 e^{\frac{1}{2}v_0 s^2} \\
\tfrac{\partial^6}{\partial s^6}\Phi(s) &= 15v_0{}^3 e^{\frac{1}{2}v_0 s^2} + 45s^2 v_0{}^4 e^{\frac{1}{2}v_0 s^2} + 15s^4 v_0{}^5 e^{\frac{1}{2}v_0 s^2} + s^6 v_0{}^6 e^{\frac{1}{2}v_0 s^2}
\end{aligned} \tag{2-29}$$

The Taylor expansion is most easily computed about $s = 0$ (the Maclaurin series for the characteristic function). This corresponds to calculating the matching probability density function about its mean, using its central moments. In this example the value of $\Phi(s)$ and its derivatives where $s = 0$ are

$$\Phi(0) = 1 \tag{2-30}$$

$$\begin{array}{lll}
\frac{\partial}{\partial s}\Phi(0) = 0 & \frac{\partial^7}{\partial s^7}\Phi(0) = 0 & \frac{\partial^{13}}{\partial s^{13}}\Phi(0) = 0 \\
\frac{\partial^2}{\partial s^2}\Phi(0) = v_0 & \frac{\partial^8}{\partial s^8}\Phi(0) = 105v_0{}^4 & \frac{\partial^{14}}{\partial s^{14}}\Phi(0) = 135135v_0{}^7 \\
\frac{\partial^3}{\partial s^3}\Phi(0) = 0 & \frac{\partial^9}{\partial s^9}\Phi(0) = 0 & \frac{\partial^{15}}{\partial s^{15}}\Phi(0) = 0 \\
\frac{\partial^4}{\partial s^4}\Phi(0) = 3v_0{}^2 & \frac{\partial^{10}}{\partial s^{10}}\Phi(0) = 945v_0{}^5 & \frac{\partial^{16}}{\partial s^{16}}\Phi(0) = 2027025v_0{}^8 \\
\frac{\partial^5}{\partial s^5}\Phi(0) = 0 & \frac{\partial^{11}}{\partial s^{11}}\Phi(0) = 0 & \\
\frac{\partial^6}{\partial s^6}\Phi(0) = 15v_0{}^3 & \frac{\partial^{12}}{\partial s^{12}}\Phi(0) = 10395v_0{}^6 &
\end{array}$$

Expanding the Maclaurin series up to the 16th degree generates the following:

$$\Phi(\omega) \approx 1 - \frac{v_0}{2!}\omega^2 + \frac{3v_0{}^2}{4!}\omega^4 - \frac{15v_0{}^3}{6!}\omega^6 + \frac{105v_0{}^4}{8!}\omega^8 - \frac{945v_0{}^5}{10!}\omega^{10} + \frac{10395v_0{}^6}{12!}\omega^{12} - \frac{135135v_0{}^7}{14!}\omega^{14} + \frac{2027025v_0{}^8}{16!}\omega^{16} \tag{2-31}$$

Figure 2.4 shows the characteristic function, and some of the approximating polynomials.

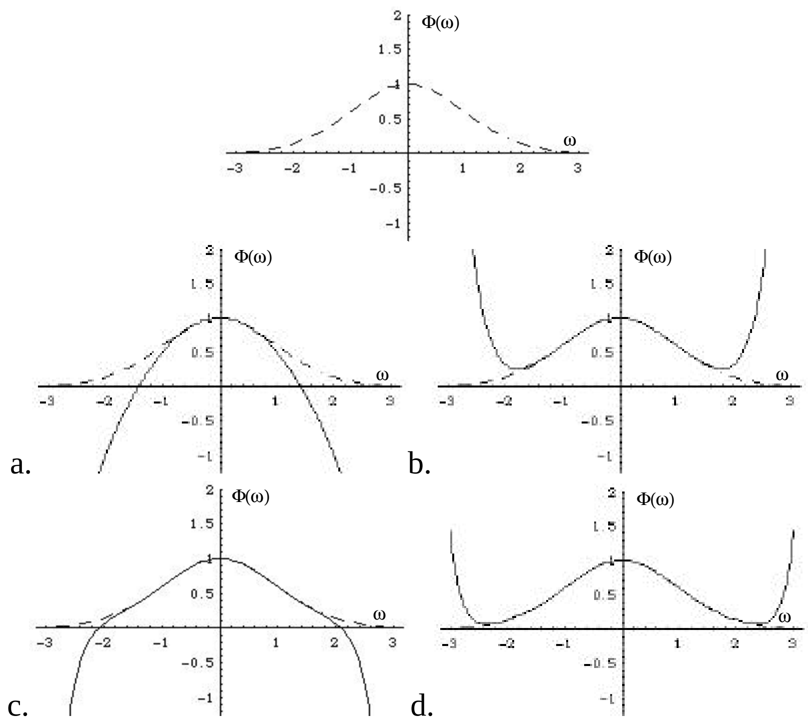

Figure 2.4. *Top:* Characteristic function for a zero mean Gaussian. Maclaurin approximating polynomials (a) n=2, (b) n=8, (c) n=10, and (d) n=16.

Bivariate Characteristic Functions

In the bivariate case the relationship between the two random variables is described by the *joint characteristic function*. Given two random variables $\tilde{u}_1$ and $\tilde{u}_2$ such that $\mathbf{u} = (\tilde{u}_1, \tilde{u}_2)$ and their joint distribution function $f(\tilde{u}_1, \tilde{u}_2)$, the joint characteristic function $F(\omega_1, \omega_2)$ and its inverse are the integrals

$$F(\omega_1, \omega_2) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(\tilde{u}_1, \tilde{u}_2) e^{i(\omega_1 \tilde{u}_1 + \omega_2 \tilde{u}_2)} d\tilde{u}_1 \, d\tilde{u}_2 \tag{2-32a}$$

$$f(\tilde{u}_1, \tilde{u}_2) = \frac{1}{4\pi^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} F(\omega_1, \omega_2) e^{-i(\omega_1 \tilde{u}_1 + \omega_2 \tilde{u}_2)} d\omega_1 \, d\omega_2 \tag{2-32b}$$

From the moment theorem it is possible to obtain $f(\tilde{u}_1, \tilde{u}_2)$ from its joint moments. Given an infinite series of central moments that describe $f(\tilde{u}_1, \tilde{u}_2)$, it is possible to the generate the probability distribution using the following Maclaurin series expansion.

$$F(\omega_1, \omega_2) = \sum_{n=0}^{\infty} \tfrac{1}{n!} \sum_{k=0}^{\infty} \binom{n}{k} \mu_{\tilde{\mathbf{u}}}^{(k, n-k)} (i\omega_1)^k (i\omega_2)^{n-k} \tag{2-33}$$

2.6. Moment Invariants of Image Functions

Hu introduced the family of moment invariants, taking advantage of the moment theorem that provides a bijection from derivatives in image space to moments in frequency [Hu 1962]. Reiss added refinements to the fundamental theorem of moment invariants almost thirty years after Hu [Reiss 1991]. A typical calculation for computing the regular moment m_{pq} of the continuously differentiable function f(x,y) is:

$$m_{pq} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^p y^q \; f(x, y) dx dy \tag{2-34}$$

As with moments of the probability density function, it is possible with a continuously differentiable function f(x,y) to generate moments m_{pq} of any order. The infinite set of moments uniquely determines f(x,y). As with moments of probability distributions, for $u \in \mathsf{R}$ and $v \in \mathsf{R}$ there exists a moment generating function $\overline{\mathrm{M}}(u, v)$ for these invariants.

$$\overline{\mathrm{M}}(u, v) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{ux+vy} \; f(x, y) dx \, dy \tag{2-35}$$

Likewise, the complete set of moments can be expanded in a power series such that

$$\overline{\mathrm{M}}(u, v) = \sum_{p=0}^{\infty} \sum_{q=0}^{\infty} m_{pq} \frac{u^p}{p!} \frac{v^q}{q!} \tag{2-36}$$

Central moments $m_{(pq)}$ are defined as

$$m_{(pq)} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - {}^{m_{10}}\!/\!_{m_{00}})^p (y - {}^{m_{01}}\!/\!_{m_{00}})^q \; f(x, y) dx \, dy \tag{2-37}$$

It is important to distinguish these invariants from statistical moments of image intensity. The central moments of image function are equivalent to moments where the spatial origin has been shifted to the image centroid $\left({}^{m_{10}}\!/\!_{m_{00}}, {}^{m_{01}}\!/\!_{m_{00}}\right)$. By contrast, moments of image intensity are centered about the mean intensity value.

To summarize, the primary difference is that moments of an image function capture image geometry while moments of image intensity describe the probability distribution of the image intensity values. Image function moments are sensitive to noise while intensity moments attempt to characterize noise. Later chapters will attempt to reimplement intensity moments to incorporate scale (Chapter 4) and image geometry (Chapter 5).